

mlguide - First concept of a machine learning toolkit

Sonja Jäckle, Rieke Alpers, Max Westphal

MOTIVATION

- Machine learning (ML) is commonly used to answer medical research questions [1] and its development might be difficult for non-experts, since there are many potential pitfalls. [2]
- In the KI-FDZ project, the application to and analysis of healthcare data in the German Health Data Lab (<https://www.healthdatalab.de/>) is redesigned. In this context, mlguide can potentially support ML practitioners.

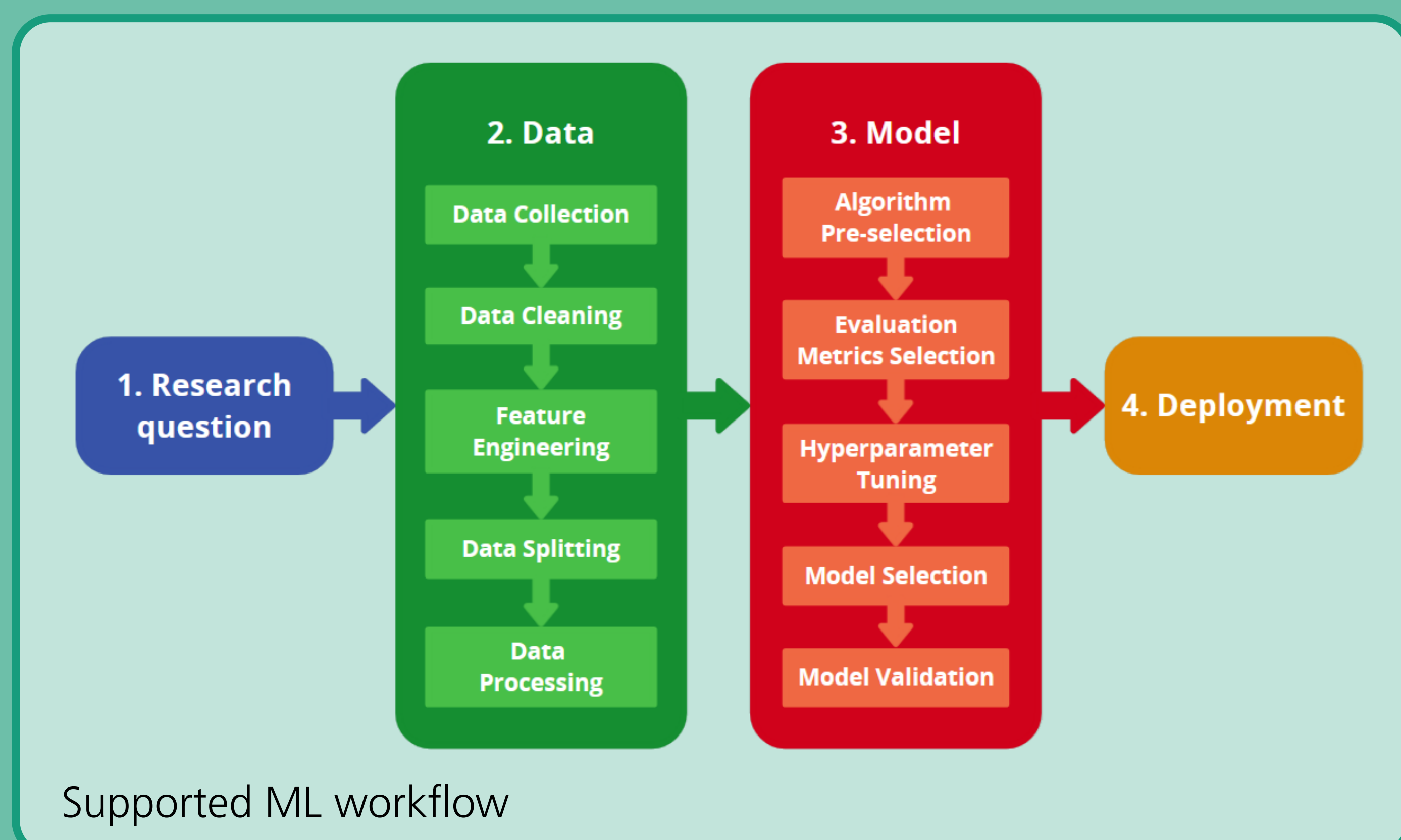
GOAL OF MLGUIDE

The aim is to offer an interactive navigation during the ML development similar to expert systems by recommending suitable methods based on current evidence [3]:

- Information about the research questions and metadata of the dataset is entered
- A recommendation for each step of the ML workflow can be requested

ML WORKFLOW

The workflow consists of four parts: problem definition – data processing – model development – deployment. Guidance support is given for data and model parts and general recommendation are given for deployment.



USER RESEARCH

- A first survey was created to determine the most urgent problems of ML development. Currently it is open and everybody who had or has any contact with ML model development is invited to take part.
- After the survey evaluation, interviews are planned to gain more specific insights where and how guidance is needed the most.

15 min survey



s.fhg.de/mlguide

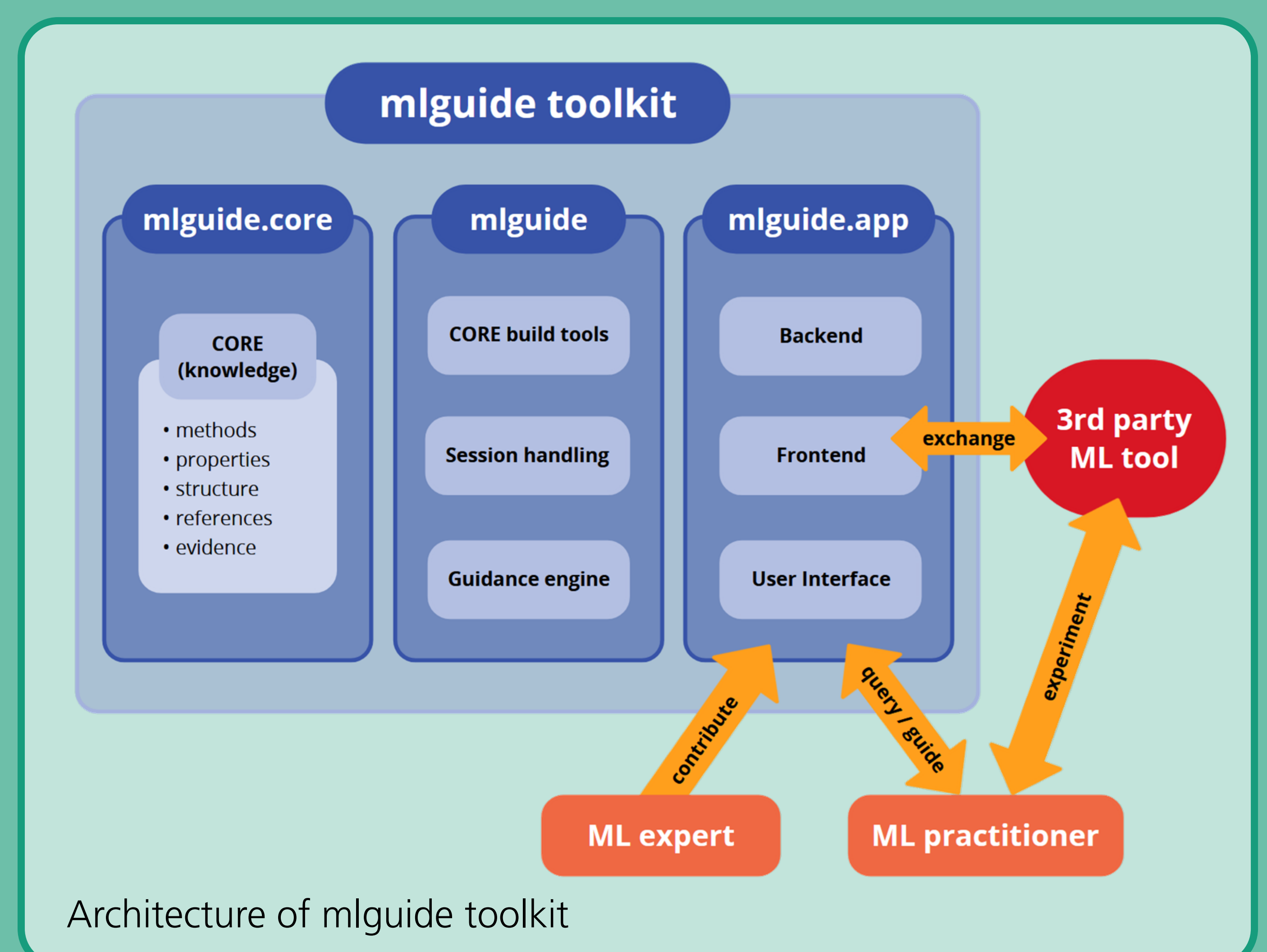
REFERENCES

- [1] Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalibi AI, Gandomi AH. Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*. 2022; 145(105458)
- [2] Haller MC et al. Prediction models for living organ transplantation are poorly developed, reported and validated: a systematic review. *Journal of Clinical Epidemiology*. 2022.
- [3] Tan CF, Wahidin LS, Khalil SN, Tamaldin N, Hu J, Rauterberg GWM. The application of expert system: A review of research and applications. *ARP Journal of Engineering and Applied Sciences* 2016;11(4):2448-2453.

MLGUIDE TOOLKIT

The toolkit consists of three parts:

- mlguide.core:
 - Collection and management of the expert knowledge and evidence.
- mlguide:
 - Organization of the guidance process by loading the available knowledge and recommending the most suitable method(s) with a guidance engine.
 - mlguide is implemented as an R package.
- mlguide.app:
 - Web application that guides the user during ML development.
 - The user can visit each step of our workflow where support is needed. Information about the research problem and the used data has to be entered for obtaining a recommendation.
 - mlguide.app is implemented with the R Shiny framework.



FUTURE WORK

- Adaption and improvements of the mlguide toolkit based on the survey result and user interviews.
- Evidence management in mlguide and automatic knowledge extraction from publication.
- Extension of the current ML workflow for unsupervised machine learning methods and unstructured data.

CONTACT

Sonja Jäckle, sonja.jaekle@mevis.fraunhofer.de

This work was funded based on a resolution of the German parliament by the German Federal Ministry of Health (BMG, Project KI-FDZ, Funding Code: 2521DAT01B).